

Active transposition in zebrafish

(transposable element/*Danio rerio*/two-dimensional gel electrophoresis/genomic transposon display)

WAN L. LAM, TIEN-SHUN LEE, AND WALTER GILBERT

Department of Molecular and Cellular Biology, The Biological Laboratories, Harvard University, 16 Divinity Avenue, Cambridge MA 02138

Contributed by Walter Gilbert, July 15, 1996

ABSTRACT We have identified a new family of Tc1-like transposons in the zebrafish, *Danio rerio*. The sequence of a candidate active transposon, deduced from sample Tzf elements, shows limited resemblance to the previously described Tdr1 elements of zebrafish. Both the Tzf and the Tdr elements are extremely abundant in zebrafish. We describe here a general strategy for detecting transposition events in a complex genome and demonstrate its utility by selectively monitoring hundreds of potentially active Tzf copies in the zebrafish genome against a background of other related elements. We have followed members of a zebrafish pedigree, using this two-dimensional transposon display strategy, to identify the first examples of active transposition of such elements in vertebrates.

Although DNA-mediated transposable elements have recently been shown to exist in a variety of vertebrates (1–13), an active copy has not been found. Such transposons typically carry a transposase gene flanked by short inverted repeats, and they are widespread in the invertebrates (14–18). The P elements of *Drosophila* and the Tc1 elements of *Caenorhabditis* are a significant source of mutations in these animals (14, 17). Vertebrate Tc1s have been identified in agnathian (jawless) and osteichthyan (bony) fishes (3–6), as well as in amphibians (7). The Tes1 element of the agnathian *Eptatretus stouti* (2), the Tdr1 elements of the zebrafish *Danio rerio* (4, 5), the SALT1, Tss, and Tsg elements of salmonoids (3, 4), and the TXr and TXz elements of *Xenopus* (7) identified to date have all suffered insertions and/or deletions and contain stop codons in their transposase gene.

Remnants (1–4, 7) of such DNA-mediated transposable elements in genes imply that they were once mobile, but demonstrating active transposition in a vertebrate genome has been an overwhelming task. Based on the transposition frequencies of the well-studied invertebrate example, the Tc1 element of *Caenorhabditis elegans* (14), one would anticipate the need to examine thousands of progeny at many chromosomal loci to detect transposition events in a single generation. We describe here a novel transposon display strategy and its use to detect active transposition of a new vertebrate transposon family in the zebrafish.

MATERIALS AND METHODS

Cloning and Sequence Analysis of Tzf Elements. Using a degenerate oligonucleotide probe corresponding to the amino acid sequence QQDNDPKHT—a conserved block in the transposases of *C. elegans* and *E. stouti*—we isolated several Tzf-containing lambda clones from a library of partially digested zebrafish genomic DNA library (Stratagene). Restriction fragments of interest were subcloned into plasmid vectors. DNA sequences of five Tzf elements were determined using

the Sequenase system (United States Biochemical) and the CircumVent system (New England Biolabs).

DNA sequences were imported into the GENETIC DATA ENVIRONMENT (19) for analysis. DNA and amino acid sequence alignments were generated by the CLUSTAL V software (20), with both fixed and floating gap penalty set at 15. Dot plot analysis of opposite DNA strands was performed, using the DOTTY PLOTTER program written by Don Gilbert (University of Indiana) to identify inverted repeats by scanning for matches of 18 in a moving window of 25 nucleotides. DNA and protein sequence data base searches were performed using various BLAST search routines (21, 22).

Two-Dimensional Transposon Display. The zebrafish used in our experiments originated from the Ekwill Fish Farms (Gibsonton, FL), and they were bred and maintained by the Harvard Fish Facility. Each frozen zebrafish specimen was individually ground into powder in liquid nitrogen, resuspended in lysis buffer (20 mM Tris-HCl, pH 7.5/5 mM Na₂EDTA/0.5% SDS), and digested with proteinase K (0.5 mg/ml) at 50°C for several hours. DNA was precipitated from the phenol-extracted lysate with 2.5 volumes of ethanol.

Zebrafish DNA was digested to completion with *Hae*III–*Hinf*I. Samples of 10–20 µg were separated on a nondenaturing acrylamide gel [acrylamide/6% *N,N'*-methylene-bis-acrylamide (wt/vol); 37.5:1] in 40 mM Tris/20 mM sodium acetate/1 mM Na₂EDTA buffer at 7 V/cm for 13.5 hr. Each gel lane (slice) was cut and placed on top of a gel containing an ascending gradient of denaturing agents [0–28% formamide (vol/vol) and 0–5.6 M urea] in the same buffer. An apparatus (ISO-DALT system; Hoeffer) facilitated simultaneous electrophoresis of multiple gels in a 25-liter buffer chamber for 16 hr at 100 V, 1.5 A at a constant temperature of 62°C. DNA was electroblotted to nylon membrane and hybridized to an oligonucleotide probe corresponding to a part of the inverted end repeat. The membranes were washed extensively (in 50 mM sodium phosphate, pH 7.2/1% SDS) at room temperature and finally at 52°C for 2 min. Autoradiographs were scanned and analyzed using VISAGE 4.6 software from (BioImage–Millipore). Twenty micrograms of a parental DNA mixture from two parents was compared with 10 µg from each diploid offspring.

RESULTS AND DISCUSSION

Tzf Elements of Zebrafish. We have identified a new Tc1-like element in zebrafish. The consensus of several homologous copies describes a 1621-bp candidate active structure, which we call Tzf (Fig. 1 *A* and *B*). This transposon resembles the recently reported zebrafish Tdr1 element (4, 5) in overall structure but shares only 38% of its amino acid residues (Fig. 1*B*). Phylogenetic analysis of vertebrate transposase groups Tzf with the SALT1 element of *Salmo solar* (3) and the TXr element of *Xenopus laevis* (7).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Data deposition: The sequences reported in this paper have been deposited in the GenBank data base (accession nos. U51226–U51230).

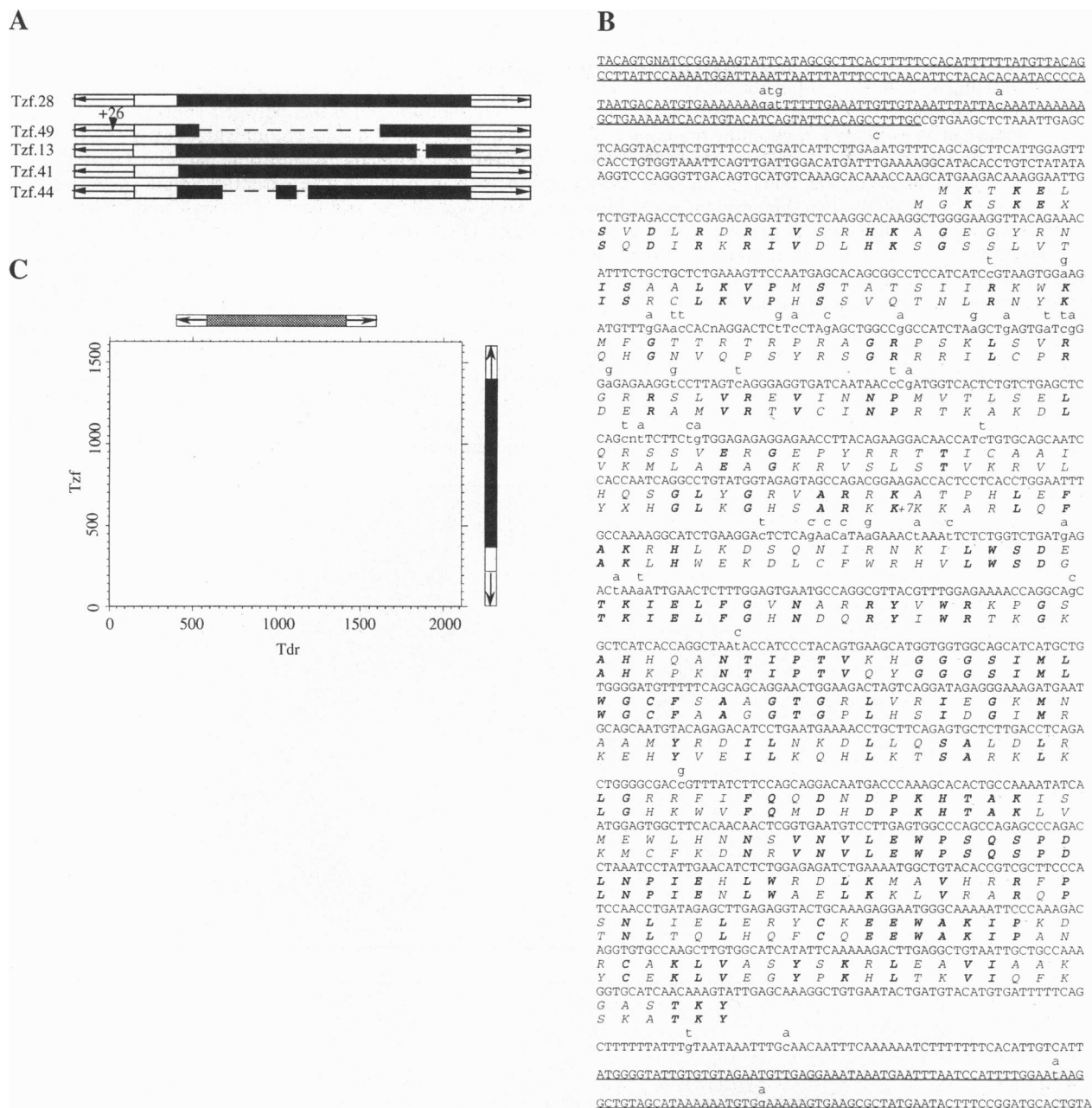


FIG. 1. Structure of Tzf elements and comparison of two zebrafish transposons. **A** 1621-nt consensus DNA sequence was derived from five Tzf sequences (GenBank accession nos. U51226–U51230). The structure of the individual Tzf elements are shown in **A**, with arrows representing the inverted repeats and black boxes representing the transposase coding regions. Major insertions (as the number of nucleotides inserted) and deletions (as gaps) are indicated. The consensus DNA sequence shown in **B** is defined by majority at any given position. At the positions where two possible nucleotides could reside, both are shown in lowercase letters along with the probable amino acid translation [that would fit the multiple vertebrate Tc1 transposase alignment shown in (7)]. The inverted repeats are underlined. The 340 amino acids of the Tdr transposase sequence (4, 5) are aligned below the 339 residues of the Tzf transposase translation with shared residues marked in bold and unidentified residues marked by an X. In **C**, the Tdr DNA sequence is plotted against the Tzf consensus, by scoring for 18 matches in a moving window of 25 nucleotides, to detect identity between their end repeats (arrows) and their transposase gene (shaded boxes). The numbers correspond to Tzf sequence positions and Tdr sequence positions (GenBank accession no. ZEFTC1).

Dot plot analysis, comparing the consensus Tzf DNA sequence against its own complementary strand, reveals a pair of 220-bp inverted repeats encompassing the deduced transposase gene. Alignment of the sequences flanking five pairs of inverted repeats identifies the presence of a TA dinucleotide at 8 of the 10 termini (presumably the other two have mutated to TC and CA), suggesting that Tzf, like Tc1 and mariner, integrates at the sequence TA. Although the originally identified Tdr element appeared to have much shorter end repeats of 52 bp (4), a dot plot of opposite strands of Tdr sequence (GenBank accession no. ZEFTC1) reveals a long extension of

the repeats after a small region of mismatches. [We also observe 200-bp inverted repeats in *X. laevis* Tc1-like transposons (7).] The Tzf and the Tdr end repeats are similar in size but very different in sequence. A dot plot of Tzf against Tdr not only shows that there is no significant sequence identity between the Tzf and the Tdr end repeats but also illustrates the low overall similarity between these two transposon types (Fig. 1C). Lowering the stringency from 18 matches to 12 matches in a moving window of 25 positions does not improve the detection of the end repeats.

Genomic Transposon Display. Since a dot plot comparison of these two transposons shows a lack of conservation of their

inverted end repeats (Fig. 1C), they can be monitored separately in the same genome using end repeat-specific probes, and, since the trans-acting Tc1A transposase of *C. elegans* functions by binding specifically within the inverted repeats (23), the number of potentially active Tzfs and Tdrs in the zebrafish genome can be estimated by counting the number of conserved terminal repeats.

To resolve the Tzf and Tdr elements in the 2-gigabase zebrafish genome, we adopted a two-dimensional gel electrophoresis system (24–26). Digestion of genomic DNA with a pair of restriction enzymes that cuts frequently within and outside the transposon but not in the inverted repeat generates two target fragments for each transposable element. We first resolve DNA fragments by size on a nondenaturing polyacrylamide gel and then separate by base sequence composition on a denaturing gradient gel in a second dimension. After transfer of the two-dimensional pattern to a nylon membrane, target fragments appear as spots upon hybridization to a single-stranded end repeat-specific probe. Fig. 2 shows a pattern of

524 hybridization signals to the Tdr probe in a wild-type zebrafish, while more than 800 signals appear with the Tzf probe. Thus, ≈ 700 Tc1-like elements are equipped with inverted repeats and are potentially active in the zebrafish genome.

We monitored Tzf loci between generations by comparing two-dimensional displays of individual offspring from a single cross to a template of the total parental DNA (father and mother, combined). The parental sample shows all the heritable Tzf alleles, and each offspring should inherit a subset. The DNA concentrations, 20 μg of parental and 10 μg of progeny, are such that a single heterozygous locus in the parent should appear at the same intensity in the progeny. We followed 51 randomly selected loci (spots) on 12 individual offspring displays to estimate the frequency of polymorphism. In strain Ekwil, 65% of the Tzf loci examined were polymorphic (33/51), while 35% (18/51) were homozygous in one or both parents (accounting for two of four, three of four, or four of four possible alleles) and appear in all offspring. Fig. 3 shows the data and our analysis to estimate how many single and

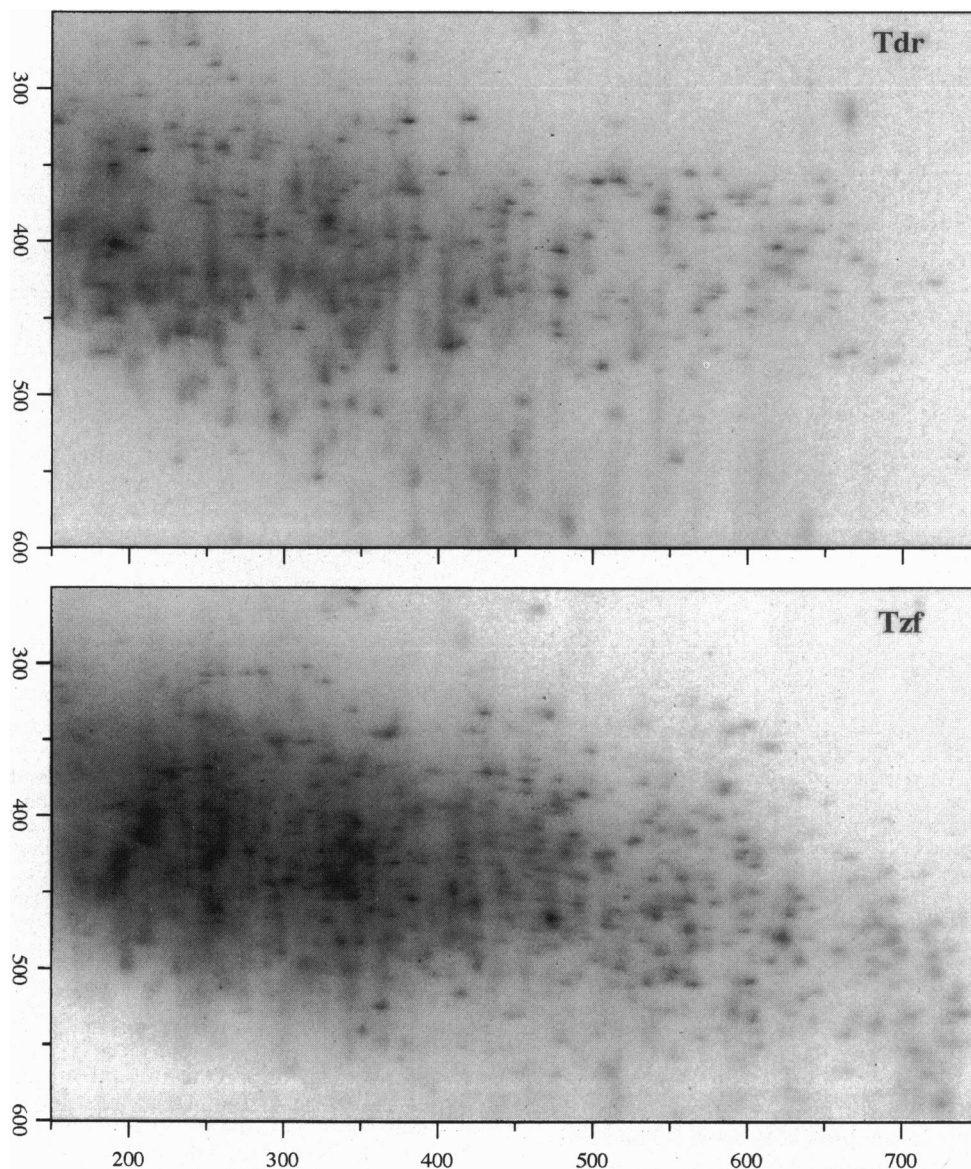


FIG. 2. Two-dimensional transposon displays. *Hae*III-*Hinf*I-digested zebrafish DNA (from one of the 25 offspring samples described in the Table 1 legend) was separated in two dimensions as described. Blotted DNA was first hybridized to a Tdr end repeat-specific probe (5'-AAATGGTTATGTGCCTTTTATACAGTGTATGTAAAC), then stripped and rehybridized with a Tzf end repeat probe (5'-AGCTGAAAAATCACATGTACATCAGTATTCACAGCCTTTG). The vertical axis indicates the nondenaturing dimension, while the horizontal axis indicates the denaturing dimension of the displayed image. Part of the whole $21 \times 19 \text{ cm}^2$ blot is displayed here as an 980×820 pixel image.

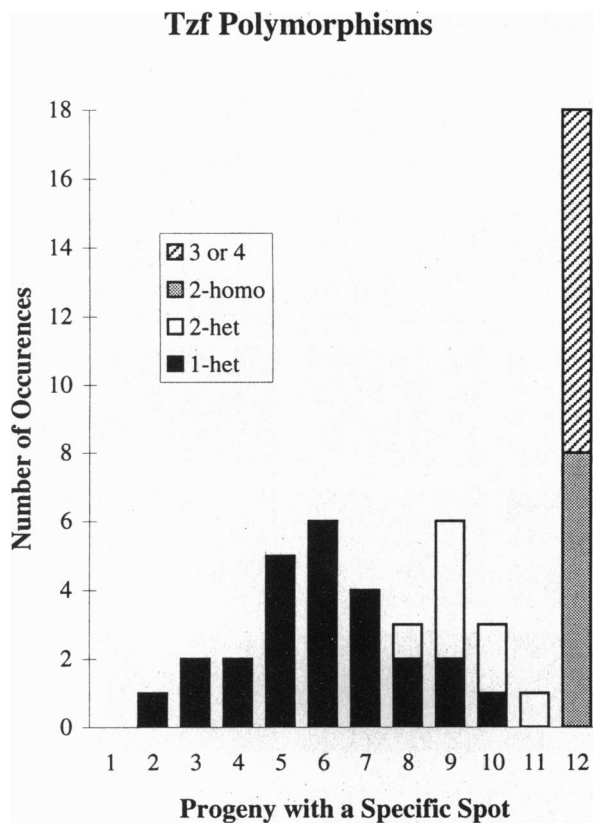


Fig. 3. Tracing Tzf loci between generations to estimate the frequency of polymorphisms. Fifty-one randomly chosen Tzf loci were monitored by comparing the two-dimensional displays of 12 offspring against the combined display of their parents, scoring for the presence of specific spots at the individual loci. The number of occurrences (number of loci out of 51 analyzed) was plotted against the frequency of progeny with a specific spot. The height of each column indicates the number of loci with that inheritance frequency. To deduce the number of single (1-het) and double (2-het) heterozygous polymorphisms from the inheritance frequency, we estimated the number of occurrences of having a specific spot present in one-half and in three-quarters of the progeny. We observed 25 ($\approx 50\%$ of all loci) and 8 ($\approx 15\%$), respectively, assuming a normal distribution with a peak at 6 (out of 12) and a peak at 9 (out of 12). The 18 spots that appear in all the progeny represent homozygosity in one or both parents. Both single homozygosity (none in the other parent) and double heterozygosity account for two of the four parental alleles. If the frequency of these two categories were the same, then 8 of the 18 universal spots would be attributed to single homozygosity (2-homo). The remaining 10 would be double homozygous, or homozygous in one and heterozygous in the other parent (3 or 4).

double polymorphisms there were. A single heterozygous polymorphism (on one of the four parental chromosomes) would be inherited by half the progeny as a spot of equal intensity. We estimate that 50% of the loci in Ekwil are of this type. A shared heterozygous polymorphism (one chromosomal locus in each parent) should appear as a spot of equal intensity in one-fourth of the progeny, of half intensity in one-half of the progeny, and be missing in one-fourth of the progeny. We estimate 15% of the loci in Ekwil are of this type. We expect an equivalent 15% of the loci to be homozygous in one parent and absent in the other and estimate that a last 20% occur in three or four copies; all of these would appear as nonpolymorphic in this test.

Active Transposition. Does active transposition occur? To assay for new insertions, we traced anonymous Tzf-containing loci from one generation to the next. We hoped that screening all the chromosomal loci at once would eliminate any need to examine thousands of offspring. We compared the Tzf display

of the total parental DNA (father and mother, mixed) with displays of 25 individual progeny from a single cross. A spot that appears in a progeny display but not in the parental sample is a candidate for a new Tzf locus.

We detected 11 new Tzf loci in the 25 offspring examined. Fig. 4 shows a number of examples. A new Tzf allele unique to a single offspring could be attributed either to a mutation at a specific restriction site or to a genuine transposition event. We were surprised to observe that ten of the novel spots are present in more than one offspring (Table 1 and Fig. 4). These ten multiple occurring new Tzf alleles (in 13–63% of the offspring) happen too frequently to be attributed to random mutations at restriction sites. We further examined the combined parental display for faint signals at these 11 novel Tzf loci, and for four cases, there are faint parental spots of varying intensities (Table 1). These do not represent polymorphisms, since all alleles on the parental blot, including single heterozygous loci, should appear at equal or higher intensity than in the progeny.

Could these novel spots be due to general genomic alterations, such as a change in methylation pattern? A methylation difference between parent and offspring, such that a sequence that was methylated in the parent becomes demethylated in the offspring and thus produces a novel fragment might be conjectured. However, the enzymes that we used, *HaeIII* and *HinfI*, are not sensitive to methylation, so this cannot be a source of the novelty. Furthermore, the spot pattern of each offspring analyzed is otherwise consistent with inheritance, and these novel spots are randomly distributed among the offspring. Another alternative interpretation would be incomplete enzymatic digestion, resulting in a fragment not being completely released from the parent, but being completely released and appearing in some of the offspring. This was controlled for by examining different exposures of several different blots of parental DNA, each blot involving an independent digestion, to see that the appearance or the lack of appearance (or faint appearance) was a consistent property, independent of the accident of digestion.

None of the 11 novel spots comes in pairs. We do not have an explanation for not having detected the expected second spot, except we are certainly seeing less than half of the spots and expect a detection bias in that we can observe the phenomenon only in reasonably clear regions of the pattern. (In the pattern of Fig. 2, we can use only the outer regions.)

Transposition of Tc1-like elements is frequently associated with the excision of the element from one site and its reintegration elsewhere (27, 28). Transposition events that occur during meiosis should appear as unique events in single offspring. Similarly, somatic transposition during mitosis in an offspring should appear as unique events. However, mitotic transposition in the parental germ line should lead to mosaicism of the parental germ cells and appear as recurring spots in a fraction of the offspring, although it is still not detectable in the total somatic parental DNA. A very early somatic event, which might be detectable as a faint spot in the parent, might influence all the germ cells and thus appear in half the progeny. (Such an event has to occur during the first few cell divisions to produce enough material for subsequent detection on a two-dimensional display—detection of 1/32 or 1/16 the intensity of a typical spot is challenging.) Indeed, in four cases, we observed a very faint signal in the parental sample (Table 1 and Fig. 4), and dark forms of these spots generally occurred in half the progeny.

Mutation Rate Contribution. Could the background mutation rate be due to active transposition? We detected a total of 87 new spots. Correcting for the number of offspring analyzed, we estimate that there were 3.9 changes/offspring; and, on correcting for the detection efficiency of $<49\%$ (see Table 1 legend), we observe a transposition frequency of about eight per individual. Suppose the zebrafish has 50,000 genes

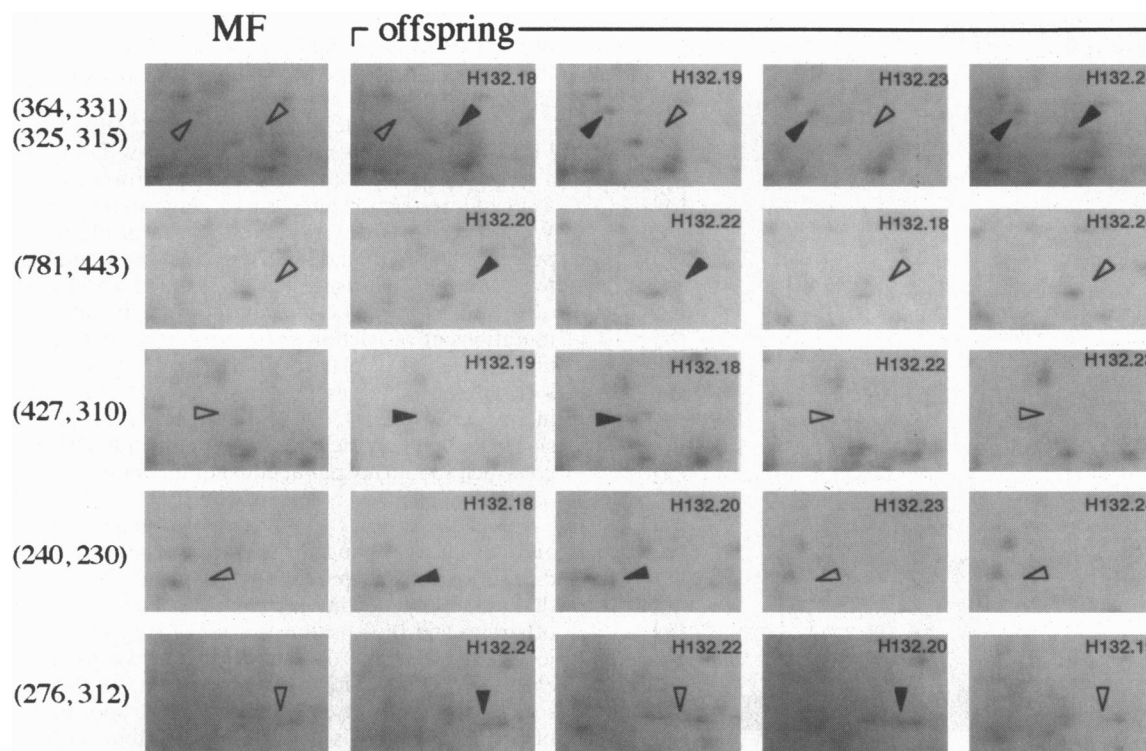


FIG. 4. Novel TzF loci in a pedigree. We followed hundreds of potentially active TzFs in wild-type (Eckwill strain) zebrafish from one generation to the next by comparing TzF displays of the 25 progeny samples to the parental ones (see Table 1 caption). The corresponding regions in a genomic TzF display of the parental DNA (combined father and mother, MF) is shown along with four progeny samples. The name of each offspring is given within the individual display frames. A black arrow points to the novel spot, while an open arrow marks the corresponding position of the absent spot or a very faint spot in the parental sample. The coordinates of these spots on the parental display correspond to those given in Table 1. Faint parental spots appear at (364, 331), (325, 315), (427, 310), and (276, 312).

Table 1. Frequencies of new TzF loci

Location of novel spot*	Mutant offspring	Offspring scored†	Mutant offspring, %	Appearance in parental template‡
240, 230	8	20	40	NV
276, 312	13	25	52	+
274, 322	10	25	40	NV
325, 315	15	25	60	+
364, 331	11	25	44	+
427, 310	8	22	36	+
514, 252	4	23	17	NV
781, 443	4	19	21	NV
508, 233	3	24	13	NV
561, 600	10	16	63	NV
689, 440	1	21	4.8	NV
Mutations/offspring§			3.9	

*Location of a novel spot is given as the corresponding coordinates on the parental template (pixels from the top left corner of the parental display, the origin).

†Twenty-five individual progeny displays were compared to three separate parental displays (each contains the total DNA from both parents) for novel spots. Any regions in a display that are not distinctly resolved are excluded from our analysis—areas near the edges of a display, regions with dense hybridization signals, and the occasional areas distorted during electrophoresis or electroblotting—are scored as unknown. Therefore, information is not available for some of the 25 offspring at certain loci. The well-resolved areas typically include about half of the TzF spots (<49%, or 400 of the >822 spots, on the parental template were analyzed); thus, approximately half of the zebrafish genome is covered.

‡A + symbol indicates the presence of a faint spot in the parental display, while NV means not visible. All the novel TzF loci with a corresponding faint parental spot are shown in Fig. 4.

§Mutations/offspring was calculated by extrapolating the observed number of mutants at each locus to 25 scoreable offspring, giving a total of 97.7 mutations per 25, or 3.9 per individual.

that occupy half of its genome. If every entry into a transcription unit were mutational, the mutation rate per gene would be 8×10^{-5} per generation, or 0.01%. This frequency is comparable to the 0.01% average background mutation rate observed by Nüsslein-Volhard and coworkers (29) for four pigmentation genes; thus, mitotic TzF transpositions in the germ lines may be responsible for a significant fraction of mutations in the zebrafish. We do not have an estimate for the contribution of meiotic transposition to the overall mutation rate—a novel spot present in multiple offspring is readily identified, whereas a singular occurrence is more prone to escape detection.

In summary, we suggest that TzF's move from one place into another in the zebrafish genome at a frequency easily detectable in a single generation. The frequency suggests that active transposition may make a significant contribution to the overall mutation rate. This fact, along with the ability to identify novel insertions in a complex genome, should lead to the development of transposon-based genetic tools for this popular model organism.

The authors wish to thank the Harvard Fish Facility for providing a family of wild-type zebrafish and C. E. Alvarez for discussion. This work was supported by grants from the National Institute of Health. W.L.L. was supported by a postdoctoral fellowship from the Medical Research Council of Canada.

- Henikoff, S. (1992) *The New Biologist* 4, 328–388.
- Heierhorst, J., Lederis, K. & Richter, D. (1992) *Proc. Natl. Acad. Sci. USA* 89, 6798–6802.
- Goodier, J. L. & Davidson, W. S. (1994) *J. Mol. Biol.* 241, 26–34.
- Radice, A. D., Bugaj, B., Fitch, D. H. A. & Emmons, S. W. (1995) *Mol. Gen. Genet.* 244, 606–612.
- Izsvak, Z., Ivics, Z. & Hackett, P. B. (1995) *Mol. Gen. Genet.* 247, 312–322.

6. Ivics, Z., Izsvak, Z., Minter, A. & Hackett, P. B. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 5008–5013.
7. Lam, W. L., Seo, P., Robison, K., Virk, S. & Gilbert, W. (1996) *J. Mol. Biol.* **257**, 359–366.
8. Morgan, G. T. (1995) *J. Mol. Biol.* **254**, 1–5.
9. Oosumi, T., Belknap, W. R. & Garlick, B. (1995) *Nature (London)* **378**, 672.
10. Smit, A. F. A. & Riggs, A. D. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 1443–1448.
11. Reiter, L. T., Murakami, T., Koeuth, T., Pentao, L., Muzny, D. M., Gibbs, R. A. & Lupski, J. R. (1996) *Nat. Genet.* **12**, 288–297.
12. Auge-Gouillou, C., Bigot, Y., Pollet, N., Hamelin, M. H., Meunier-Rotival, M. & Periquet, G. (1995) *FEBS Lett.* **368**, 541–546.
13. Robertson, H. M., Zumpano, K. L., Lohe, A. R. & Hartl, D. L. (1996) *Nat. Genet.* **12**, 360–361.
14. Moerman, D. G. & Waterston, R. H. (1989) in *Mobile DNA*, eds. Howe, M. & Berg, D. (Am. Soc. Microbiol., Washington, DC), pp. 537–556.
15. Capy, P., David, J. R. & Hartl, D. L. (1992) *Genetica* **86**, 37–46.
16. Robertson, H. M. (1993) *Nature (London)* **362**, 241–245.
17. Robertson, H. M. (1995) *J. Insect Physiol.* **41**, 99–105.
18. Robertson, H. M. & Lampe, D. J. (1995) *Annu. Rev. Entomol.* **40**, 333–357.
19. Smith, S. W., Overbeek, R., Woese, C. R., Gilbert, W. & Gillevet, P. M. (1994) *Comput. Appl. Biosci.* **10**, 671–675.
20. Higgins, D. G. & Sharp, P. M. (1988) *Gene* **73**, 237–244.
21. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403–410.
22. Altschul, S. F., Boguski, M. S., Gish, W. & Wootton, J. C. (1994) *Nat. Genet.* **6**, 119–129.
23. Van Luenen, H. G. A. M., Colloms, S. D. & Plasterk, R. H. A. (1993) *EMBO J.* **12**, 2513–2520.
24. Fischer, S. G. & Lerman, L. S. (1979) *Cell* **16**, 191–200.
25. Uitterlinden, A. G., Slagboom, P. E., Knook, D. L. & Vijg, J. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 2742–2746.
26. Sidman, C. L. & Shaffer, D. J. (1994) *Genomics* **23**, 15–22.
27. Vos, J. C., Van Luenen, H. G. & Plasterk, R. H. (1993) *Genes Dev.* **7**, 1244–1253.
28. Van Luenen, H. G. A. M., Colloms, S. D. & Plasterk, R. H. A. (1994) *Cell* **79**, 293–301.
29. Mullins, M. C., Hammerschmidt, M., Haffter, P. & Nüsslein-Volhard, C. (1994) *Curr. Biol.* **4**, 189–202.